

# Appendix B Terminology and Glossary

- [Acronyms and Abbreviations](#)
- [Terminology](#)

## Acronyms and Abbreviations

<b>CCSDS</b>	Consultative Committee for Space Data Systems
<b>CMIS</b>	Content Management Interoperability Services
<b>CERIF</b>	Common European Research Information Format
<b>DDS</b>	Data Distribution Service for Real-Time Systems
<b>ENVRI</b>	Environmental Research Infrastructure
<b>ENVRI_RM</b>	ENVRI Reference Model
<b>ESFRI</b>	European Strategy Forum on Research Infrastructures
<b>ESFRI-ENV RI</b>	ESFRI Environmental Research Infrastructure
<b>GIS</b>	Geographic Information System
<b>IEC</b>	International Electrotechnical Commission
<b>ISO</b>	International Organisation for Standardization
<b>OAIS</b>	Open Archival Information System
<b>OASIS</b>	Organization for the Advancement of Structured Information Standards
<b>ODP</b>	Open Distributed Processing
<b>OGC</b>	Open Geospatial Consortium
<b>OMG</b>	Object Management Group
<b>ORCHESTRA</b>	Open Architecture and Spatial Data Infrastructure for Risk Management
<b>ORM</b>	OGC Reference Model
<b>OSI</b>	Open Systems Interconnection
<b>OWL</b>	Web Ontology language
<b>SOA</b>	Service Oriented Architecture
<b>SOA-RM</b>	Reference Model for Service Oriented Architecture
<b>RDF</b>	Resource Description Framework
<b>RM-OA</b>	Reference Model for the ORCHESTRA Architecture
<b>RM-ODP</b>	Reference Model of Open Distributed Processing
<b>UML</b>	Unified Modelling Language
<b>W3C</b>	World Wide Web Consortium
<b>UML4ODP</b>	Unified Modelling Language For Open Distributed Processing

## Terminology

**Access Control:** A functionality that approves or disapproves of access requests based on specified access policies.

**Acquisition Service:** Oversight service for integrated data acquisition.

**Active role:** A active role is typically associated with a human actor.

**Add Metadata:** Add additional information according to a predefined schema (metadata schema). This partially overlaps with data annotations.

**Annotate Data:** Annotate data with meaning (concepts of predefined local or global conceptual models).

**Annotate Metadata:** Link metadata with meaning (concepts of predefined local or global conceptual models). This can be done by adding a pointer to concepts within a conceptual model to the data. If e.g. concepts are terms in and SKOS/RDF thesaurus, published as linked data then this would mean entering the URL of the term describing the meaning of the data.

**Annotation:** (verb) The action of annotating or making notes. (noun) A note added to anything written, by way of explanation or comment.

**Annotation Service:** Oversight service for adding and updating records attached to curated datasets.

**Assign Unique Identifier:** Obtain a unique identifier and associate it to the data.

**Authentication:** A functionality that verifies a credential of a user.

**Authentication Service:** Security service responsible for the authentication of external agents making requests of infrastructure services.

**Authorisation:** A functionality that specifies access rights to resources.

**Authorisation Service:** Security service responsible for the authorisation of all requests made of infrastructure services by external agents.

**Backup:** A copy of (persistent) data so it may be used to restore the original after a data loss event.

**Behaviour :** A behaviour of a community is a composition of actions performed by roles normally addressing separate business requirements.

**Build Conceptual Models:** Establish a local or global model of interrelated concepts.

**Capacity Manager:** An active role, which is a person who manage and ensure that the IT capacity meets current and future business requirements in a cost-effective manner.

**Carry out Backup:** Replicate data to an additional data storage so it may be used to restore the original after a data loss event. A special type of backup is a long term preservation.

**Catalogue service:** Oversight service for cataloguing curated datasets.

**Check Quality:** Actions to verify the quality of data.

**Citation:** from the ENVRI RM perspective, citation is defined as a pointer from a publication to:

- data source(s)
- and/or the owner(s) of the data source(s)
- a description of the evaluation process, if available
- a timestamp marking the access time to the data sources, thus reflecting a certain version

**Citizen (synonyms: General Public, Media):** An active role, a person, who is interested in understanding the knowledge delivered by an environmental science research infrastructure, or discovering and exploring the [Knowledge\\_Base\\_Glossary](#) enabled by the research infrastructure.

**Citizen Scientist:** An active role, member of the general public who engages in scientific work, often in collaboration with or under the direction of professional scientists and scientific institutions (also known as amateur scientist).

**Community:** A collaboration which consists of a set of roles agreeing their objective to achieve a stated business purpose.

**Concept:** Name and definition of the meaning of a thing (abstract or real thing). Human readable definition by sentences, machine readable definition by relations to other concepts (machine readable sentences). It can also be meant for the smallest entity of a conceptual model. It can be part of a flat list of concepts, a hierarchical list of concepts, a hierarchical thesaurus or an ontology.

**Conceptual Model:** A collection of concepts, their attributes and their relations. It can be unstructured or structured (e.g. glossary, thesaurus, ontology). Usually the description of a concept and/or a relation defines the concept in a human readable form. Concepts within ontologies and their relations can be seen as machine readable sentences. Those sentences can be used to establish a self-description. It is, however, practice today, to have both, the human readable description and the machine readable description. In this sense a conceptual model can also be seen as a collection of human and machine readable sentences. Conceptual models can reside within the persistence layer of a data provider or a community or outside. Conceptual models can be fused with the data (e.g. within a network of triple stores) or kept separately.

**Coordination Service:** An oversight service for data processing tasks deployed on infrastructure execution resources.

**Data Acquisition Community.** A community, which collects raw data and bring (streams of) measures into a system.

**Data Acquisition Subsystem:** A subsystem that collects raw data and brings the measures or data streams into a computational system.

**Data Analysis:** A functionality that inspects, cleans, transforms data, and provides data models with the goal of highlighting useful information, suggesting conclusions, and supporting decision making.

**Data Assimilation:** A functionality that combines observational data with output from a numerical model to produce an optimal estimate of the evolving state of the system.

**Data Broker:** Broker for facilitating data access/upload requests.

**Data Cataloguing:** A functionality that associates a data object with one or more metadata objects which contain data descriptions.

**Data Citation:** A functionality that assigns an accurate, consistent and standardised reference to a data object, which can be cited in scientific publications.

**Data Collection:** A behaviour performed by a *Data Collector* that control and monitor the collection of the digital values from a *sensor* instrument or a human sensor, such as a *Measurer* or a *Observer*, associating consistent time-stamps and necessary metadata.

**Data Collector:** Active or passive role, adopted by a person or an instrument collecting data.

**Data Consumer:** Either an active or passive role, which is an entity who receives and use the data.

**Data Curation Community:** A community, which curates the scientific data, maintains and archives them, and produces various data products with metadata.

**Data Curation Subsystem:** A subsystem that facilitates quality control and preservation of scientific data.

**Data Curator:** An active role, which is a person who verifies the quality of the data, preserve and maintain the data as a resource, and prepares various required data products.

**Data Discovery & Access:** A functionality that retrieves requested data from a data resource by using suitable search technology.

**Data Exporter:** Binding object for exporting curated datasets.

**Data Extraction:** A functionality that retrieves data out of (unstructured) data sources, including web pages, emails, documents, PDFs, scanned text, mainframe reports, and spool files.

**Data Identification:** A functionality that assigns (global) unique identifiers to data contents.

**Data Importer:** An Oversight service for the import of new data into the data curation subsystem.

**Data infrastructure:** a collection of data assets, organisations that operate and maintain them and guides describing how to use and manage the data. A data infrastructure is sustainably funded and has oversight that provides direction to maximise data use and value by meeting the needs of society. Data infrastructure includes technology, processes and organisation.

**Data management:** a process development and execution of architectures, policies, practices and procedures in order to manage the data lifecycle needs of a specific research community.

**Data management plan (DMP):** a formal document that outlines how data are to be handled both during a research project and after the project is completed.

**Data Mining:** A functionality that supports the discovery of patterns in large data sets.

**Data Originator:** Either an active or a passive role, which provide the digital material to be made available for public access.

**Data Processing Control:** A functionality that initiates the calculation and manages the outputs to be returned to the client.

**Data Processing Subsystem:** A subsystem that aggregates the data from various resources and provides computational capabilities and capacities for conducting data analysis and scientific experiments.

**Data Product Generation:** A functionality that processes data against requirement specifications and standardised formats and descriptions.

**Data Provenance:** Information that traces the origins of data and records all state changes of data during their lifecycle and their movements between storages.

**Data Provider:** Either an active or a passive role, which is an entity providing the data to be used.

**Data Publication:** A functionality that provides clean, well-annotated, anonymity-preserving datasets in a suitable format, and by following specified data-publication and sharing policies to make the datasets publically accessible or to those who agree to certain conditions of use, and to individuals who meet certain professional criteria.

**Data Publication Community:** A community that assists the data publication, discovery and access.

**(Data Publication) Repository:** A passive role, which is a facility for the deposition of published data.

**Data Publishing Subsystem:** A subsystem that enables discovery and retrieval of data housed in data resources.

**Data Quality Checking:** A functionality that detects and corrects (or removes) corrupt, inconsistent or inaccurate records from data sets.

**Data Service Provision Community:** A community that provides various services, applications and software/tools to link, and recombine data and information in order to derive knowledge.

**Data State:** Term used as defined in ISO/IEC 10746-2. At a given instant in time, data state is the condition of an object that determines the set of all sequences of actions (or traces) in which the object can participate.

**Data Storage & Preservation:** A functionality that deposits (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and makes them accessible on request.

**Data Store Controller:** A data store within the data curation subsystem.

**Data Transfer Service:** Oversight service for the transfer of data into and out of the data curation subsystem.

**Data Transmission:** A functionality that transfers data over communication channel using specified network protocols.

**Data Transporter:** Generic binding object for data transfer interactions.

**Data Use Community:** A community who makes use of the data and service products, and transfers the knowledge into understanding.

**Data Use Subsystem:** A subsystem that provides functionalities to manage, control, and track users' activities and supports users to conduct their roles in the community.

**Describe Service:** Describe the accessibility of a service or process, which is available for reuse, the interfaces, the description of behaviour and/or implemented algorithms.

**Design of Measurement Model:** A behaviour that designs the measurement or monitoring model based on scientific requirements.

**Do Data Mining:** Execute a sequence of metadata / data request --> interpret result --> do a new request

**e-Infrastructure:** a combination and interworking of digitally-based technology (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to support modern, internationally leading collaborative research be it in the arts and humanities or the sciences.

**Educator (synonym: Trainer):** An active role, which is a person who makes use of the data and application services for education and training purposes.

**Engineer (synonym: Technologist):** An active role, which is a person who develops and maintains the research infrastructure.

**Environmental Scientist:** An active role, which is a person who conduct research or perform investigation for the purpose of identifying, abating, or eliminating sources of pollutants or hazards that affect either the environment or the health of the population. Using knowledge of various scientific disciplines, may collect, synthesize, study, report, and recommend action based on data derived from measurements or observations of air, food, soil, water, and other sources.

**ENVRI Reference Model:** A common ontological framework and standards for the description and characterisation of computational and storage systems of ESFRI environmental research infrastructures.

**Experiment Laboratory:** Community proxy for conducting experiments within a research infrastructure.

**Field Laboratory:** Community proxy for interacting with data acquisition instruments.

**Final review:** Review the data to be published, which will not likely be changed again.

**Free text annotation:** to add a short explanation or opinion to a text or drawing (equivalent to the dictionary definition of annotation).

**Instrument Controller:** An integrated raw data source.

**Knowledge Base:** (1) A store of information or data that is available to draw on. (2) The underlying set of facts, assumptions, and rules which a computer system has available to solve a problem.

**Knowledge infrastructure:** robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds.

**Mapping Rule:** Configuration directives used for model-to-model transformation.

**(Measurement Model) Designer:** An active role, which is a person who design the measurements and monitoring models based on the requirements of environmental scientists.

**Measurement result:** Quantitative determinations of magnitude, dimension and uncertainty to the outputs of observation instruments, sensors (including human observers) and sensor networks.

**Measurer:** An active role, which is a person who determines the ratio of a physical quantity, such as a length, time, temperature etc., to a unit of measurement, such as the meter, second or degree Celsius.

**Metadata:** Data about data, in scientific applications is used to describe, explain, locate, or make it easier to retrieve, use, or manage an information resource.

**Metadata Catalogue:** A collection of metadata, usually established to make the metadata available to a community. A metadata catalogue has an access service.

**Metadata Harvesting (Publishing Community Role):** A behaviour performed by a metadata harvester to gather metadata from data objects in order to construct catalogues of the available information. A functionality that (regularly) collects metadata (in agreed formats) from different sources.

**Metadata Harvester (Publishing Community Role):** A passive role performed by a system or service collecting metadata to support the construction /selection of a global conceptual model and the production of mapping rules.

**Metadata State:**

- raw: are established metadata, which are not yet registered. In general, they are not shareable in this status
- registered: are metadata which are inserted into a metadata catalogue.
- published: are metadata made available to the public, the outside world. Within some metadata catalogues registered.

**Passive Role:** A passive role is typically associated with a non-human actor.

**Perform Mapping:** Execute transformation rules for values (mapping from one unit to another unit) or translation rules for concepts (translating the meaning from one conceptual model to another conceptual model, e.g. translating code lists).

**Persistent Data:** Term (data) used as defined in ISO/IEC 10746-2. Data is the representations of information dealt by information systems and users thereof. Data which are persistent (stored).

**Perform Measurement or Observation:** Measure parameter(s) or observe an event. The performance of a measurement or observation produces measurement results.

**PID Generator:** A passive role, a system which assigns persist global unique identifiers to a (set of) digital object.

**PID Registry:** A passive role, which is an information system for registering PIDs.

**PID Service:** External service for persistent identifier assignment and resolution.

**Policy Maker (synonym: Decision Maker):** An active role, a person, who makes decisions based on the data evidences.

**Process Control:** A functionality that receives input status, applies a set of logic statements or control algorithms, and generates a set of analogue / digital outputs to change the logic states of devices.

**Process Controller:** Part of the execution platform provided by the data processing subsystem.

**Process Data:** Process data for the purposes of:

- converting and generating data products
- calculations: e.g., statistical processes, simulation models
- visualisation: e.g., alpha-numerically, graphically, geographically

Data processes should be recorded as provenance.

**Provenance:** The pathway of data generation from raw data to the actual state of data.

**Publish Data:** Make data public accessible.

**Publish Metadata:** Make the registered metadata available to the public.

**QA Notation:** Notation of the result of a Quality Assessment. This notation can be a nominal value out of a classification system up to a comprehensive (machine readable) description of the whole QA process.

**Quality Assessment (QA):** Assessment of details of the data generation, including the check of the plausibility of the data. Usually the quality assessment is done by predefined checks on data and their generation process.

**Query Data:** Send a request to a data store to retrieve required data.

**Query Metadata:** Send a request to metadata resources to retrieve metadata of interests.

**Observer:** An active role, which is a person who receives knowledge of the outside world through the senses, or records data using scientific instruments.

**Raw Data Collector:** Binding object for raw data collection.

**Reference Mode:** A reference mode is an abstract framework for understanding significant relationships among the entities of some environment.

**Register Metadata:** Enter the metadata into a metadata catalogue.

**Research Infrastructure:** means facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling Information and Communications Technology-based infrastructures such as Grid, computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research. Such infrastructures may be "single-sited" or "distributed" (an organised network of resources) [41].

**Resource Registration:** A functionality that creates an entry in a resource registry and inserts resource object or a reference to a resource object in specified representations and semantics.

**Role :** A role in a community is a prescribing behaviour that can be performed any number of times concurrently or successively.

**Science Gateway:** Community portal for interacting with an infrastructure.

**Scientific Modelling and Simulation:** A functionality that supports the generation of abstract, conceptual, graphical or mathematical models, and to run an instance of the model.

**Scientist (synonym: Researcher):** An active role, which is a person who makes use of the data and application services to conduct scientific research.

**(Scientific) Workflow Enactment:** A specialisation of Workflow Enactment, which support of composition and execution a series of computational or data manipulation steps, or a workflow, in a scientific application. Important processes should be recorded for provenance purposes.

**Security Service:** Oversight service for authentication and authorisation of user requests to the infrastructure.

**Semantic Annotation:** link from an information object (single datum, data set, data container) to a concept within a conceptual model, enabling the discovery of the meaning of the information object by human and machines.

**Semantic Broker:** Broker for establishing semantic links between concepts and bridging queries between semantic domains.

**SV Community Behaviour:** A behaviour enabled by a *Semantic Mediator* that unifies similar data (knowledge) models based on the consensus of collaborative domain experts to achieve better data (knowledge) reuse and semantic interoperability.

**Semantic Laboratory:** Community proxy for interacting with semantic models.

**Semantic Mediator:** A passive role, which is a system or middleware facilitating semantic mapping discovery and integration of heterogeneous data.

**Sensor:** A passive role, which is a converter that measures a physical quantity and converts it into a signal which can be read by an observer or by an (electronic) instrument.

**Sensor Network:** A passive role, which is a network consists of distributed autonomous sensors to monitor physical or environmental conditions.

**Service:** Service or process, available for reuse.

**Service Consumer:** Either an active or a passive role, which is an entity using the services provided.

**Service Description:** Services and processes, which are available for reuse, be it within an enterprise architecture, within a research infrastructure or within an open network like the Internet, shall be described to help avoid wrong usage. Usually such descriptions include the accessibility of the service, the description of the interfaces, the description of behavior and/or implemented algorithms. Such descriptions are usually done along service description standards (e.g. WSDL, web service description language). Within some service description languages, semantic descriptions of the services and/or interfaces are possible (e.g. SAWSDL, Semantic Annotations for WSDL)

**Service Provider:** Either an active or a passive role, which is an entity providing the services to be used.

**Service Registry:** A passive role, which is an information system for registering services.

**Setup Mapping Rules:** Specify the mapping rules of data and/or concepts.

**Specification of Investigation Design:** This is the background information needed to understand the overall goal of the measurement or observation. It could be the sampling design of observation stations, the network design, the description of the setup parameters (interval of measurements) and so on... It usually contains important information for the allowed evaluations of data. (E.g. the question whether a sampling design was done randomly or by strategy determines which statistical methods that can be applied or not).

**Specification of Measurements or Observations:** The description of the scientific measurement model which specifies:

- what is measured;
- how it is measured;
- by whom it is measured; and
- what the temporal design is (single /multiple measurements / interval of measurement etc. )

**Specify Investigation Design:** specify design of investigation, including sampling design:

- geographical position of measurement or observation (site) -- the selections of observations and measurement sites, e.g., can be statistical or stratified by domain knowledge;
- characteristics of site;
- - preconditions of measurements.

**Specify Measurement or Observation:** Specify the details of the method of observations/measurements.

**Stakeholder (synonyms: Private Investor, Private Consultant ):** An active role, a person, who makes use of the data and application service for predicting market so as to make business decision on producing related commercial products.

**Storage:** A passive role, which is memory, components, devices and media that retain digital [computer data](#) used for computing for some interval of time.

**Storage Administrator:** An active role, which is a person who has the responsibilities to the design of data storage, tune queries, perform backup and recovery operations, raid mirrored arrays, making sure drive space is available for the network.

**Store Data:** Archive or preserve data in persistent manner to ensure continuing accessible and usable.

**Subsystem:** a set of capabilities that collectively are defined by a set of interfaces with corresponding operations that can be invoked by other subsystems. Subsystems can be executed independently, and developed and managed incrementally.

**Technician:** An active role, which is a person who develop and deploy the sensor instruments, establishing and testing the sensor network, operating, maintaining, monitoring and repairing the observatory hardware.

**Track Provenance:** Add information about the actions and the data state changes as data provenances.

**Unique Identifier (UID):** With reference to a given (possibly implicit) set of objects, a unique identifier (UID) is any identifier which is guaranteed to be unique among all identifiers used for those objects and for a specific purpose.

**User Behaviour Tracking:** A behaviour enabled by a Data Use Subsystem that to track the Users. User Behaviour Tracking is the analysis of visitor behaviour on a website. The analysis of an individual visitor's behaviour may be used to provide options or content that relates to their implied preferences; either during a visit or in the future visits. Additionally, it can be user to track content use and performance.

**User Group Work Supporting:** A behaviour enabled by a Data Use Subsystem that to support controlled sharing, collaborative work and publication of results, with persistent and externally citable PIDs.

**User Profile Management:** A behaviour enabled by a Data Use Subsystem that to support persistent and mobile profiles, where profiles will include preferred interaction settings, preferred computational resource settings, and so on.

**User Working Space Management:** A behaviour enabled by a Data Use Subsystem that to support work spaces that allow data, document and code continuity between connection sessions and accessible from multiple sites or mobile smart devices.

**User Working Relationships Management:** A behaviour enabled by a Data Use Subsystem that to support a record of working relationships, (virtual) group memberships and friends.

**Virtual Laboratory:** Community proxy for interacting with infrastructure subsystems.

**Virtual Research Environment (VRE, synonyms: Science Gateway, Collaboratory, Digital Library, Inhabited Information Space, Virtual Laboratory):** a web-based working environment tailored to serve the needs of a research community. A VRE is expected to provide an array of commodities needed to accomplish the research community's goal(s); it is open and flexible with respect to the overall service offering and lifetime; and it promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution.